A STUDY OF INTER AND INTRA EXAMINER RELIABILITY IN MARKING ESSAYS WITH AND WITHOUT USING THE MARKING SCHEME

PRINCIPAL INVESTIGATOR
DR. BIPIN J. JHAVERI
M. A , Ph. D.

RESEARCH ASSOCIATE
BABUBHAI C. PATEL
M. Sc., M. Ed.

NEW ARTS COLLEGE
VALLABH VIDYANAGAR
DISTRICT : KAIRA

P R E F A C E

It is a pleasure for me to put before our educationists the results of a project undertaken by my College under a grant from the National Council of Educational Research and Training, Ministry of Education, Government of India.

Some of the hypotheses have not stood the test of this research. Still I have faith in them. I suggest the project be again taken up by some institution to recheck them with a bigger sample and an examiner's meeting before the assessment, for which there was no provision in this project.

I heartily thank the N.C.E.R.T., The Chairman, The Secretary and the Joint Secretary of the Charutar Vidya Mandal. Dr. R.H. Dave, Dr. Prafulla Patel, the Principal M.B. Patel College of Education and his colleagues Shri Babubhai C. Patel and Shri Haribhai G. Desai and last but not the least the Vice-Chancellor of Sardar Patel University and the Examiners who undertook the most boring task of going through the same essays a number of times, and with an extremely hair-splitted marking scheme.

BIPIN J. JHAVERI
Principal Investigator
New Arts College
Vallabh Vidyanagar

-1-1968

C O N T E N T S

CHAPTER I

INTRODUCTION

Most of the examinations, conducted by the Universities or the public bddies, include essay-type questions. One of the major defects of such examinations is that the subjectivity of the examiner makes the judgement unreliable. Personal views, beliefs, attitudes and, some times even whims of the examiners are reflected to a great extent in the marking of the essay-type questions.

The low correlations between the assessments made by the different examiners is the other defect of the essay-type questions. Many a times the correlation between the two assessment, of the same question made by same examiner at two different times, is also found to be very discouraging. Objectivity and reliability of the assessment are reduced to the minimum in case of language subjects.

We cannot do away with the essay-type questions in language subjects. Essays-type questions constitute a major tool for assessing the student's power of written expression. Richness of language and variety of the vocabulary are reflected well only in the essay-type examinations. The essay-type questions, therefore, are very essential in the examination of language subjects.

The task of retaining the essay-type questions in the examination, and at the same time increasing the objectivity as well as the reliability of the assessment is indeed very difficult. However we must find out ways and means to achieve this objective. The investigator, inspired by this problem, thought of conducting a research project to find out ways and means for increasing the inter-examiner and intra-examiner reliability in marking of the essays-type questions, particularly in assessing the essays asked in the language question paper.

There are many aspects which decrease objectivity and reliability of the assessment of essays. One of the main factors is that the headings of the essays contain very few words or some times, at the most, a phrase or two. This does not bring home to the examinee or the examiner what is expected of him. Consequently the examinee may write some thing different from what the examiner expects; and the paper setter might have an altogether different thing in his mind while setting the paper. Thus the sole aim of asking the question might be refuted. A complete liberty regarding the content and the scope of the question is given to the examinee, some times leading to most fantastic result.

The second major factor is that none of three

persons, the paper setter, the examiner and the examinee, knows the weightage to be given to each of the different aspects of the written expression such as the language, grammar, presentation, literary value, quotations, clarity of thought, lucidity, paragraphing, etc. Hence it is very likely that the weightages given by different examiners to the various aspects of the written expression may differ to a great extent.

To avoid these two difficulties, regarding clarity of the meaning of a particular essay and the weightages to be given to each aspect of the written expression, two ways are suggested. Firstly, of the essays are well defined and the specific points of importance are made clear to the examiner and the examinee, the two will not differ in their interpretation of the content and the scope of the essays. Secondly, if a good marking scheme with specifications regarding the weightages to be given to each of the various aspects of the written expression is prepared, the difficulty in assigning proper weightages to the different objectives could be overcome. It was expected that, these two modifications would increase objectivity and reliability of the assessment.

The Problem :

With this view in mind the following title of the

research project was formulated :

"A study of inter-examiner and intra-examiner reliability in marking the traditional as well as the well-defined essays, with and without the use of marking scheme."

Here the traditional essay means an essay which is commonly asked in routine. The well-defined essay is one which has a heading which includes the points to be stressed. They are specified and the meaning of the essay is made more clear, e.g. "Your visit to an imaginary city which consists of, among other things, a river, a road, a castle, a temple and a museum." In the former essay the examinee may write any thing he likes while in the later case he is expected to describe nature, place of culture, places of historical interest, idols and deities, etc.

Hypothesis to be tested :

With the views expressed in the above few paragraphs in mind, the investigator formulated the following four hypotheses for the purpose of the research project.

1. Within each type of essays, the intra-examiner reliability of an examiner in the case of marking

essays with the help of marking scheme will be significantly higher than that in the case of marking essays without the help of a marking scheme.

2. With and without the help of marking scheme the intra-examiner reliability of an examiner in the case of well-defined essays will be significantly higher than that in the case of traditional essays.

3. Within each type of essays, the average inter-examiner reliability in the case of marking essays with the help of marking scheme will be significantly higher than that in the case of marking essays without the help of a marking scheme.

4. The average inter-examiner reliability in the case of well defined essays will be significantly higher than that in the case of traditional essays in the case of both with an without the help of a marking scheme.

CHAPTER II

PROCEDURE OF THE INVESTIGATION

The first step of the investigation was to construct a test form consisting of different categories of essays. This was done on the basis of a work-paper devised by the investigator which itself was based on the experience gained at S.S.C. Examination Paper-setters' and Moderators' work-shop held at Ahmedabad in the month of December 1965. The specimens of the final test forms alongwith their translation in English are given in Appendix - A.

The next step in the procedure was to list the different abilities required for effective essay-writing at the level of Pre-University Examination. The other thing was to decide the weightages to be given to each of the various abilities required in the effective written expression. The two things, (1) listing of the abilities and (2) deciding the weightage for each of the abilities were completed in the work-shop organised under the joint auspices of N.C.E.R.T. and S.S.C. Examination Board, Gujarat State, under guidance of Dr. R.H. Dave, the present Head, Department of Curriculum and Evaluation, N.C.E.R.T., New Delhi.

The participant-language teachers of Gujarati in the camp fixed by consensus the various abilities required. The same group decided the weightage to be given to each ability. The following table shows the distribution of marks according to the weightages given to different abilities of the written expression.

TABLE I

Knowledge Ability pertaining to	Weightage in marks
1. Content matter	6
2. Logical Presentation	6
3. Quality of Language	4
4. Correctness of Language	4
Total	20

Each one of the four different aspects was analysed in detail and sectional weightages were decided for important points of each aspect. (The detailed marking scheme is given in appendix C). Both the test forms were developed in Gujarati, one consisting of traditional essays and the other consisting of well-defined essays.

A random sample, of 100 students studying in the Preparatory Arts Classes was selected for the purpose of administration of both the test forms. The sample included the students from both the courses. The "with English Course" and "without English Course". Both the test forms were administered to the random sample of students at one and the same time. Out of the 100 answer books of the students the irrelavent 17 were discarded. They were irrelevant because either they were blank or blank in one of the tests or were awfully brief, revealing insincerity of effort on part of the assess. The remaining 83 answer-books were retained for the purpose of analysis.

Assessment of the answer-books :

To carry out assessment of the answer-books, four examiners were appointed. All the answer-books of traditional essays were assessed by all the four examiners in rotation, without using the marking scheme. Then after an interval of a minimum of 20 days all the answer-books were given code numbers and were assessed by the same four examiners again without using the marking scheme. The same procedure was repeated while assessing the answer-books with the help of the marking scheme with different sets of dummy numbers for the first and second assessment. Thus every answer-books of the traditional essay was assessed

four times (twice without using the making scheme and twice with the help of marking scheme* by each of the four examiners, i.e. in all a single answer-book was assessed sixteen times. Thesame procedure was adopted for assessing answer-books of the well-defined essays.

Care was taken in distributing the answer-books in such a way that each examiner received the same answer-book after an interval of about twenty days or more for the next assessment. That is, an interval of atleast twenty days was kept in between all the four successive assessments made by each examiner, in case of each type of essays. The following table represents proposed schedule of the asse sment to be carried out by the examiners. The schedule given in the Table II was fairly adhered to, though not entirely. The time gap was, however, maintained.

TABLE II

A SCHEDULE FOR ASSESSMENT OF ANSWER-BOOKS BY DIFFERENT EXAMINERS

Month	Week of the months	Examiner A	Examiner B	Examiner C	Examiner D
First	I	*Til	-	-	-
	II	Wil	Til	-	-
	III	-	Wil	Til	-
	IV	-	-	Wil	Til

contd...

Second	I	-	-	-	Wi1
	II	Ti2	-	-	-
	III	Wi2	Ti2	-	-
	IV	-	Wi2	Ti2	-
Third	I	-	-	Wi2	Ti2
	II	-	-	-	Wi2
	III	TJ1	-	-	-
	IV	WJ1	TJ1	-	-
Fourth	I	-	WJ1	TJ1	-
	II	-	-	WJ1	TJ1
	III	-	-	-	WJ1
	IV	TJ2	-	-	-
Fifth	I	WJ2	TJ2	-	-
	II	-	WJ2	TJ2	-
	III	-	-	WJ2	TJ2
	IV	-	-	-	WJ2

* Meanings of Codes used in the above Table : T - Traditional essay. W - well-defined essay, i - without using marking scheme, J - with the marking scheme. 1 - first assessment, 2 - second assessment.

A special type of perforated sheet was attached over every answer-book for writing a dummy number on it, at different times to successive assessments. A copy of the front page of the answer-book is put as Appendix B. A file was maintained for each type of assessment and the confidential list consisting of code numbers (or the dummy numbers) along with the original numbers of the answer-books was kept with a responsible clerk. Each mark-sheet obtained from the examiner had the dummy numbers on it, so at the time of compilation of marks, the corresponding original numbers were written on every mark sheet by the person in charge of assigning the dummy numbers on the answer-books.

Each pupil had eight scores from each of the four examiners as follows:-

1. Two scores of traditional essay without the the marking scheme.

2. Two scores of traditional essay with the marking scheme.

3. Two scores of well-defined essay without the marking scheme.

4. Two scores of well-defined essay with the marking scheme.

The whole data thus obtained have been analysed by means of suitable statistical procedures and the interpretations of the analysis have been given in the following chapter. The four hypotheses have also been tested critically by applying the tests of significance of difference between the Z transformations of the different Correlation Coefficients.

* * *

C H A P T E R III

RESULTS AND DISCUSSION

Here the comparative views regarding the different statistics like mean, standard deviations and the correlation coefficients are given. The views are based on the application of rigrous statistical tests of significance of the differences between the statistics.

To save space in the tables, code letters are used for different types of assessments. The following table shows the specifications regarding the code letters used in different tables as well as in some of the paragraphs of the report.

TABLE III

Specifications Regarding Code Letters

Sr.No.	Specifications	Code letters used
1.	Traditional Essays	T
2.	Well-defined Essays	W
3.	Assessment without the marking scheme	i
4.	Assessment with the marking scheme	j
5.	First assessment	1
6.	Second assessment	2

contd...

Table IIII contd..

Sr.No.	Specifications	Code letters used
7.	Assessment by Dr. Jhaveri B.J.	A
8.	Assessment by Shri Master A.M.	B
9.	Assessment by Shri Mansuri F.J.	C
10.	Assessment by Shri Patel B.C.	D

For example : "WjD2" means the information regarding the assessment of well defined essays assessed with the marking scheme by Shri B.C. Patel for the second time. Similarly we can interpret the meaning of other combinations of the code letters.

The following table gives a summary of the analysis of the data.

TABLE IV

Comparative view of the analysis of the data

Examiner		Ti1	Ti2	Wi1	Wi2	TJ1	TJ2	WJ1	WJ2
A	Mean	6.31	7.84	5.59	6.0	9.11	8.18	8.64	7.3
	Range	0-12	0-16	1-15	1-13	1-15	3-12	1-15	0-13
	S.D.	2.1	3.21	2.48	2.76	2.06	2.22	2.38	2.20
Intra Examiner r			.7036		.8603		.7528		.6928

contd...

Table IV contd...

B	Mean	7.39	7.01	5.96	6.45	6.84	7.05	6.71	5.46
	Range	0-12	0-11	0-10	0-13	1-10	1-10	0-11	0-9
	S.D.	2.1	2.16	2.29	2.59	1.639	1.823	2.26	2.08
	Intra Examiner r	.7156		.8556		.8372		.8052	
C	Mean	7.47	7.44	0-72	7-04	9-96	9-96	9-95	9-325
	Range	0-14	1-14	1-11	1-11	3-15	3-14	1-14	1-14
	S.D.	2.538	2.571	2.1	2.1	2.77	2.33	2.697	2.329
	Intra Examiner r	.8246		.8581		.8571		.7563	
D	Mean	8.32	8.25	6.84	7.76	9.16	9.21	8.02	7.096
	Range	1-12	1-12	0-12	0-13	2-12	2-13	1-13	1-12
	S.D.	2.38	2.57	2.25	2.80	2.23	2.44	2.73	2.82
	Intra-Examiner r	.8104		.9017		.8135		.8805	

SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MEAN ASSESSMENTS:

TABLE V

Significance of difference between the means of Ti1 and TJ1 assessment

Examiner	Mean Mi of Assessment Ti1	Mean Mi of Assessment TJ1	Difference between the means D-MJ1-Mi1	Standard error of the difference D	Critical ratio t, D/ D
A	6.31	8.11	+ 1.80	0.236 (x)	7.6 *
B	7.39	6.84	- 0.55	0.179 (x)	3.073 *
C	7.47	9.98	+ 2.51	0.41 (Y)	6.136 *
D	8.32	9.16	+ 0.84	0.2261(X)	3.676 *

TABLE VI

Significance of difference between the means of assessments Wi1 and WJ1

Examiner	Mean Mi of assessment Wi1	Mean MJ of assessment WJ1	Difference between the means D=MJ-Mi	Standard error of the difference D	Critical ratio, t = D/ D
A	5.59	0.64	+ 3.05	0.377 (Y)	9.09 *
B	5.96	6.71	+ 0.75	0.1402 (X)	5.3495 *
C	6.72	9.95	+ 3.23	0.375 (Y)	8.6133
D	6.84	8.02	+ 1.18	0.389 (Y)	3.033 *

* Significant at .01 level of significance

(X) Standard error of the difference between the means found out by the formula $D = \sqrt{{}_{M1}^{2} + {}_{M2}^{2} - 2r\,M1\,M2}$

which takes into account the correlation between the ~~mm~~ score whose means are compared.

(Y) Standard error of the difference between the means found out by the formula $D = \sqrt{{}_{M1}^{2} + {}_{M2}^{2}}$

which does not take into account the correlation between the scores whose means are compared.

If the difference between two means is found to be significant by using the formula $D = \sqrt{{}_{M1}^{2} + {}_{M2}^{2}}$ for finding out the standard error of the difference between the means, it (the difference) would have been highly (much more) significant if the formula $D = \sqrt{{}_{M1}^{2} + {}_{M2}^{2} - 2r\,M1\,M2}$ were used for finding out the standard error of the difference between the means. So to save unnecessary labour of computing the coefficients of correlations of scores whose means are to be compared, the formula $D = \sqrt{{}_{M1}^{2} + {}_{M2}^{2}}$ is used where the

differences between the means were sufficiently large.

From the observations of tables XXI and XXII, it is seen that the means of the assessments of both the types of essays, carried out by using the marking scheme, are significantly higher than those of the assessments carried out without using the marking scheme, in case of the three examiners A, C and D. In case of the examiner B this is true only regarding the well-defined essays while reverse is the case regarding the traditional essays. This may be due to poor discriminative judgement on the part of examiner B. Thus in general it can be concluded that due to the marking scheme the mean of the assessment increases.

The examiner is at liberty to assign any weightage to any point, whether relevant or irrelevant, but when the marking scheme is supplied to the examiner, he is compelled to assign due weightage to very small but relevant points which might have been ignored if the marking scheme were not used. This might be the likely reason of increase in the mean assessment when the marking scheme is used.

Variations regarding the standard deviations are not consistant in any direction (i.e. increase or decrease), and hence no generalisation can be made regarding the standard deviations.

INTRA-EXAMINER RELIABILITY i.e. Testing the hypothesis 1, viz. Within each type of essays, the intra-examiner reliability of an examiner in the case of marking essays with the help of marking scheme will be significantly higher than that in the case of marking essays without the help of a marking scheme.

TABLE VII

Coefficients of correlations between the assessments Ti1 & Ti2 and TJ1 & TJ2

Examiner	rTi1,Ti2	rTJ1,TJ2	ZTi1, Ti2	ZTJ1, TJ2
A	0.7036	0.7528	0.8744	0.9796
B	0.7156	0.8372	.8986	1.2122
C	0.8246	0.8541	1.1708	1.2824
D	0.8204	0.8135	1.1282	1.1383
Average	0.7692*	0.8187*	1.0180	1.1531

* The average rTi1, Ti2 and rTJ1 TJ2 are found out on the basis of the corresponding Z transformations. All the correlation coefficients are highly significant.

TABLE VIII

Significance of different between ZTi1 Ti2 & ZTJ1, TJ2

Examiner	Difference Dz = ZTJ - ZTi	Standard error of the difference Dz	Critical Ratio t = Dz/Dz	Level of significance
A	+ 0.1052	0.1581	0.6654	not significance
B	+ 0.3136	0.1581	1.9836	.05
C	+ 0.1116	0.1581	0.7059	not significance
D	+ 0.0101	0.1581	0.0639	not significance
Average	+ 0.1351	0.1581	0.8545	not significant

As we are interested in the increase in the Intra-Examiner reliability, when the marking scheme is used, one tailed test of significance of the difference between the Z transformations is used. The tables VII and VIII reveal that to the tradditional essays, the intra-examiner reliability in the case of the assessment when the marking scheme is used, is greater than that of the assessment when the marking scheme is not used. But increase in reliability due to the use of marking scheme is not statistically significant expect in the case of examiner B. So it is not very reasonable to conclude that the marking scheme increases the intra-examiner reliability.

At the same time we cannot deny the possibility of increase in the reliability as the differences in all the cases, though not significant, are in favour of the marking scheme. So it is desirable to repeat the experiment by taking a bigger sample of about 400 students so as to get the more valid results.

TABLE IX

Intra-Examiner Coefficients of correlation between the assessments Wi1 & Wi2 and WJ1, & WJ2

Examiner	rWi, Wi2	rWJ1, WJ2	ZWi1, Wi2	ZWJ1, WJ2
A	0.8603	0.6928	1.2942	0.8533
B	0.8556	0.8052	1.2760	1.1136
C	0.8581	0.7563	1.2863	0.9876
D	0.9017	0.8805	1.4813	1.3783
Average	0.8704*	0.7944*	1.3345	1.0832

* Average rwi1, wi2 and rWJ1, WJ2 are found out on the basis of corresponding Z transformations. All the coefficients of correlation are highly significant.

TABLE X

Significance of the difference between Zwi1, wi2 & ZWJ1, WJ2

Examiner	Difference Dz = ZWJ-ZWi	Standard error of the difference Dz	Critical Ratio t= DZ/Dz	Level of significance
A	- 0.4409	0.1581	2.7887	0.01
B	- 0.1630	0.1581	1.0309	not significant
C	-0.2987	0.1581	1.8893	not significant
D	- 0.1030	0.1581	0.6515	not significant
Average	- 0.2513	0.1581	1.5895	not significant

From the above two tables it is clearly seen that the results are quite opposite to expectation. For the well-defined essays the intra-examiner reliability in case of assessment when the marking scheme is used is less than that of the assessment when the marking scheme is not used. Except in case of examiner A, all the differences, though in favour of not using the marking scheme, are not statistically significant. So it is neither reasonable to conclude that the use of marking scheme reduced the reliability, nor the possibility of decrease in reliability can be denied.

The results of traditional essays and well-defined essays are quite opposite and the differences in reliability in both the cases are not significant. And hence the hypothesis 1 cannot be safely retained.

INTRA-EXAMINER RELIABILITY i.e. Testing the hypothesis 2, viz. With and without the help of marking scheme the intra-examiner reliability of an examiner in the case of well-defined essays will be significantly higher than that in the case of traditional essays.

TABLE XI

Intra-Examiner Coefficients of correlations between the assessments Ti1 and Ti2 as well as Wi1 and Wi2.

Examiner	rTi1, Ti2	rWi1, Wi2	ZTi1,Ti2	ZWi1, Wi2
A	0.7036	0.8603	0.8744	1.2942
B	0.7156	0.8556	0.8986	1.2766
C	0.8246	0.8581	1.1708	1.2863
D	0.8104	0.9017	1.1282	1.4813
Average	0.7692*	0.8704*	1.0180	1.3346

* Average rTi1, Ti2 and rWi1, Wi2 are found out on the basis of respective average Z transformations. All the correlations coefficients are highly significant and the intra-examiner reliability coefficients in case of well defined essays are greater than those in case of traditional essays. The following table shows level of significance of the difference between rTi1, Ti2 and rWi1, Wi2.

TABLE XII

Significance of the differences between ZTi1, Ti2 & ZWi1, Wi2

Examiner	Difference Dz = ZWi-ZTi	Standard error of the difference Dz	Critical ratio t = Dz/Dz	Level of significance
A	0.4198	0.1581	2.655	.01
B	0.3780	0.1581	2.391	.01
C	0.1155	0.1581	0.724	not signfficant
D	0.3531	0.1581	2.246	.05
Average	0.3166	1.1581	2.003	.05

For finding out the level of significance of the differences between the Z transformations of the intra-examiner reliability coefficients, one tailed test is applied. The difference is not found to be signific-ant in case of examiner C. In case of other three examiners and on the average the differences are significant and so it is reasonable to conclude that the intra-examiner reliability in case of well defined essays, is greater than that in case of traditional essays, when the marking scheme is not used. The level of significance of the difference between the intra-examiner reliability coefficients, in favour of well-defined essays, is not so high when the

marking scheme is used. The following two tables give the exact picture regarding the differences between the intra-examiner reliability when the marking scheme is used.

TABLE XIII

Intra-Examiner Coefficients of correlations between the assessments TJ1 & TJ2 and WJ1 & WJ2

Examiner	rTJ1, TJ2	rWJ1,WJ2	ZTJ1,TJ2	ZWJ1,WJ2
A	0.7528	0.6928	0.9796	0.8533
B	0.8372	0.8052	1.2122	1.1136
C	0.8571	0.7563	1.2824	0.9876
D	0.8135	0.8805	1.1383	1.3783
Average	0.8187	0.794	1.1531	1.0832

The above table shows that the intra-examiner reliability coefficients are lower in case of well defined essays than in case of traditional essays for all the examiners except examiner D, when the marking scheme is used.

TABLE XIV

Significance of the difference between ZTJ1, TJ2 and ZWJ1, WJ2

Examiner	Difference Dz = ZWJ-ZTJ	Standard error of the difference Dz	Critical ratio t-Dz/Dz	Level of significance
A	- 0.1263	0.1581	- 0.7988	not significant
B	- 0.0986	0.1581	- 0.6236	not significant
C	- 0.2948	0.1581	- 1.8646	not signficant
D	+ 0.2400	0.1581	+ 1.5180	not significant
Average	- 0.0699	0.1581	- 0.4421	not significant

The above table reveals a very strange thing. Except in case of examiner D the differences, though not significant, are in favour of traditional essays when the marking scheme has no effect on positive side of increasing the intra-examiner reliability. On the contrary use of the marking scheme has washed out the significant difference in reliability which was in favour of well-defined essays when the marking scheme was not used. So nothing can be said regarding the effect of marking scheme, at the most it can be said that there seems to be no real difference in the intra-examiner reliability coefficients of the assessments TJ and WJ.

Thus only the second part of the second hypothesis can be safely retained but the first part of it cannot be retained. And so the it can be said that second hypothesis is true only in case when the marking scheme is not used. The retained part of the hypothesis can be written as

"The intra-examiner reliability in case of well defined essays is greater than that in case of traditional essays when the marking scheme is not used. "

INTER-EXAMINER RELIABILITY i.e. Testing the hypothesis 3. Viz. Within each type of essays, the average inter-examiner reliability in the case of marking essays with the help of marking scheme will be significantly higher than that in the case of marking essays without the help of a marking scheme.

contd.. Table XV

TABLE XV

Inter-Examiner coefficients of correlations for the assessments Ti1 and TJ1

Between Examiners	Inter-Examiner Coefficient of correlation rTi1	Inter-Examiner coefficient of correlation rTJ1	ZTi1	ZTJ1
A and B	0.6465	0.4546	0.7693	0.4905
A and C	0.4946	0.3320	0.5421	0.3451
A and D	0.6234	0.4775	0.7305	0.5198
B and C	0.6353	0.5732	0.7512	0.6523
C and D	0.4435	0.5422	0.4766	0.6072
Average	0.5755*	0.5096*	0.6576	0.5621

* Average rTi1 and RtJ1 are computed on the banks of corresponding z transformations. All the reliability coefficients are highly significant.

TABLE XVI

Significance of the difference between ZTi1 and ZTJ1

Between Examiners	Difference Dz = ZTJ1, ZTJ1	Standard error of difference Dz	Critical Ratio t=Dz/Dz	Level of significance
A and B	- 0.2788	.1581	- 1.7634	not significant
A and C	- 0.1970	.2581	- 1.2460	not significant
A and C	- 0.2107	.1581	- 1.3327	not significant
B and C	- 0.0989	.1581	- 0.6255	not significant
B and D	+ 0.3156	.1581	+ 1.9962	.05
C and D	+ 0.13-6	.1581	+ 0.8260	not significant
Average	- 0.0955	.1581	- 0.6040	not significant

From the above two tables it is seen that the use of marking scheme does not increase the inter-examiner reliability. Except only in two cases, all the other reliability coefficients of TJ1 are smaller than those of Ti1. Except one, all the differences between the reliability coefficients are not significant and so the first part of the hypothesis 3 cannot be safely retained. And hence it cannot be concluded that inter-examiner reliability coefficients in case of assessment TJ1 are greater than those in case of assessment Ti1.

TABLE XVII

Inter-examiner coefficients of correlations for the assessments Wi1 and WJ1

Between the Examiners	Inter Examiner Correlation rWi1	Inter Examiner Correlation rWJ1	ZWi1	ZWJ1
A and B	0.7102	0.6294	0.8876	0.7404
A and C	0.9878	0.5642	0.8437	0.6390
A and D	0.9112	0.6127	1.1306	0.7132
B and C	0.7387	0.7833	0.9476	1.0534
B and D	0.9412	0.8329	1.7515	1.1977
C and D	0.7978	0.7936	1.0925	1.0810
Average	0.8036	0.7278*	1.1089	0.9041

* The average rWi1 and rWJ1 are computed on the basis of the corresponding Z transformations. All the correlation coefficients are highly significant.

TABLE XVIII

Significance of the difference between ZWi1 and ZWJ1

Between the Examiner	Difference Dz =ZWJ-ZWi	Standard error of difference Dz	Critical Ratio t= Dz/Dz	Level of significance
A and B	- 0.1472	0.1581	- 0.9311	not significant
A and C	- 0.2047	0.1581	- 1.2948	not significant
A and D	- 0.4174	0.1581	- 2.6401	.01
B and C	+ 0.1058	0.1581	+ 0.6692	not significant
B and D	- 0.5538	0.1581	- 3.5028	.01
C and D	- 0.0115	0.1581	- 0.0727	not significant
Average	- 0.2048	0.1581	- 1.2953	not significant

From the above two tables it is clear that the differences in reliability coefficients are not in favour of using the marking scheme. Not a single difference in favour of using the marking scheme is significant. On the contrary two are significant in favour of not using the marking scheme. This is any way very surprising. And hence the hypothesis 3 ("Within each type of essay the average inter-examiner reliability in case of marking essays with the help of marking scheme will be significantly higher than that in the case of marking essays without the help of a marking scheme") is refuted.

INTER-EXAMINER RELIABILITY i.e. Testing the hypothesis 4, viz.

The average inter-examiner reliability in the case of well defined essays will be significantly higher than that in the case of traditional essays in the case of both with and without the help of a marking scheme.

TABLE XIX

Inter-Examiner Coefficients of correlation for the assessments Ti1 and Wi1

Between Examiners	Inter Examiner correlation rTi1	Inter Examiner correlation rWi1	ZTi1	ZWi1
A and B	0.6465	0.7102	0.7693	0.8876
A and C	0.4946	0.6878	0.5421	0.8437
A and D	0.6234	0.8112	0.7305	1.1306
B and C	0.6353	0.7387	0.7512	0.9476
B and D	0.5818	0.9412	0.6652	1.7515
C and D	0.4435	0.7978	0.4766	1.0925
Average	0.5755*	0.8036*	0.6558	1.1089

* The average correlations rTi1 and rWi1 are computed on the basis of the corresponding transformations. All the coefficients of correlation are highly significant.

TABLE XX

Significance of the difference between ZTi and ZWi

Between the Examiners	Difference Dz = ZWi-ZTi	Standard error of difference Dz	Critical Ratio t=Dz/Dz	Level of significance
A and B	0.1183	0.1581	0.7483	not significant
A and C	0.3016	0.1581	1.9077	.05
Aaand D	0.4001	0.1581	2.5631	.01
B and C	0.1964	0.1581	1.2422	not significant
B and D.	1.0863	0.1581	6.8709	.01
C and D	0.6159	0.1581	3.8956	.01
Average	0.4513	0.1581	2.8545	.01

From the table XX it is clearly seen that when the marking scheme is not used, the inter-examiner reliability coefficients in case of well defined essays are greater than those in case of traditional essays. Except two, all thedifferences are statistically significant and so it is reasonable to conclude that when the marking scheme is not used, well-defined essays have greater inter-examiner reliability than the traditional essays have.

TABLE XXI

Inter-Examiner coefficients of correlation for the assessment TAT1 & WJ

Between the Examiners	Inter-Examiner correlation rTJ1	Inter-Examiner correlation rWJ1	ZTJ1	ZWJ1
A and B	0.4546	0.6294	0.4905	0.7404
A and C	0.3320	0.5642	0.3451	0.6390
A and D	0.4775	0.6127	0.5198	0.7132
B and C	0.5732	0.7833	0.6623	1.0534
B and D	0.7534	0.8329	0.9808	1.1977
C and D	0.5422	0.7936	0.6072	1.0810
Average	0.5096*	0.7278*	0.5993	0.9041

* The average correlations rTJ1 & rWJ1 are computed on the basis of the corresponding Z transformations.

All the correlation coefficients are highly significant.

TABLE XXII

Significance of the difference btween ZTJ1 and ZWJ1

Between the Examiners	Difference Dz = ZWJ-ZTJ	Standard error of difference Dz	Critical Ratio t=Dz/Dz	Level of significance
A and B	0.2501	0.1581	1.5819	not significant
A and C	0.2939	0.1581	1.8595	.05
A and D	0.1934	0.1581	1.2226	not significant
B and C	0.4021	0.1581	2.5433	.01
B and D	0.2169	0.1581	1.3719	not significant
C and D	0.4738	0.1581	2.9968	.01
Average	0.3048	0.1581	1.928	.05

In the above two tables it is seen that all the inter-examiner reliability coefficients for the assessment of well defined essays are greater than those for the traditional essays, but three of them are not statistically significant. The average inter-examiner reliability coefficient for well defined essays is significantly greater than that for the traditional essay. And so it is reasonable to conclude that, when the marking scheme is

used the inter examiner reliability is greater in case of well-defined essays than that in the case of traditional essays.

On the evidences of the conclusions regarding the data of tables XX and XXII, it is fairly reasonable to retain the hypothesis 4. ("The average inter-examiner reliability in case of well defined essays will be significantly higher than that in case of traditional essays in case of both with and without the help of marking scheme.")

All the data and discussion presented in this chapter reveal the following three facts:

1. The use of the marking scheme significantly increases the mean assessment.

2. The use of the marking scheme does not necessarily increase any type of reliability.

3. The reliability of the assessment of well defined essays is greater than that of traditional essays.

A FEW OBSERVATIONS

Variations within the Examiners:

Under this sub heading only observations and not the conclusions based on the test of significance of difference between the averages, are given. During the course of analysis of the data it was found that majority of the students do not get the same number of marks at the time of second assessment by the same examiner. Some students get more marks and some get less marks in the second assessment. This may be due to the subjective nature of the content of the essay. The average differences (not mean difference) between the first and second assessment by various examiners are given in the following table:

TABLE XXIII

Average differences between the marks of the first and second assessments of traditional as well as well-defined essays

Types of Assessment	TiA	TiB	TiC	TiD	WiA	WiB	WiC	WiD
Average difference	2.1	0.8	1.0	0.8	1.0	0.77	0.82	1.45
Type of Assessment	TJA	TJB	TJC	TJD	WJA	WJB	WJC	WJD
Average difference	1.2	0.8	1.6	1.02	1.74	1.5	1.5	1.2

In six out of eight cases the average differences between the first and second assessment with the marking scheme are greater than the corresponding differences when the marking scheme was not used. This shows that marking scheme at least in this investigation did not prove useful in reducing the range of differences between the first and second assessments. When the marking scheme is not used, the average differences between the first and the second assessment of well-defined essays are less than those of the traditional essays. However this is not true in case of examiner D. Thus well defined essays are better than the traditional essays.

Variations between the Examiners :

Under this sub heading also, only observations and not conclusions based on tests of significance of differences between the averages are given. Different examiners give different number of marks while assessing the same answer-book. The following table gives the average of the maximum difference between the assessments carried out by the four examiners

TABLE XXIV

Average of the maximum differences between the marks given by different examiners

Type of Assessment	Ti	TJ	Wi	WJ
Average of Maximum differences between the marks	3.81	4.78	3.31	4.65

The differences in case of assessment without using the marking scheme are less than the corresponding those in case of assessment with using the marking scheme. This shows that at least in this investigation the marking scheme does not help in decreasing the variations between the examiners.

The differences in case of assessment of well defined essays are less than the corresponding differences in case of assessment of the traditional essays. Thus well defined essays are better then the traditional essays.

Well-defined Essays Vs. Traditional Essays :

In comparison with traditional essays, well defined essays are more clear and specific and extent of vaguness is reduced to minimum as a result intra-examiner and

inter-examiner reliability increases. Discrimination between a dull and a cleaver student become sharper in the assessment of well defined essays.

However the above observations are not found consistant when the marking scheme is used. This may be because of faulty and vague nature of the marking scheme. Vaguness in the written expfession of the students also decreases in case of well defined essay.

Use of Marking Scheme :

Results regarding marking scheme are found very discouraging. Findings of the investigations are not found in favour of the formulated hypotheses; in some cases they are found to be quite opposite to the expected ones. Marking scheme is not found to be useful in increasing any type of reliability or in decreasing the disparity of marks within and between the examiners. The use of marking scheme increases the mean assessment, because the examiners are compelled to assign due weightage to any small but important point which might be ignored by them when the marking scheme is not supplied.

Reactions of the Examiners :

During the course of analysis of the data the examiners were interviewed and their views and reactions

regarding the introduction of well defined essays and the use of marking scheme were noted down.

During the course of discussion it appeared that the marking scheme given to the examiners was desirable but not practicable. The scheme was not foolproof. There were some loop-holes regarding the weightage to be given to spellings, correctness of the language etc. Some of the aspects of the scheme were not defined in terms of specific behavioural changes in the written expression of the students and thus there was some vagueness in the marking scheme.

Two examiners were of the opinion that too much splitting of the different aspects of the written expression is not desirable, one was not able to decide either way, while the fourth one was in favour of splitting the different aspects to a great extent; as shown in Appendix C.

It was boring to all the examiners to assess the answer-books according to the marking scheme as they had to read the answer book again and again and it took too much of their time in assessing a single answer-book. Still, if results, improve, boredom should be tolerated, and made less unwelcome by higher remuneration.

C H A P T E R IV

CONCLUSIONS AND IMPLICATIONS

Conclusions :

If the results of this project be taken as valid, it could be concluded in a nutshell, as follows :

Reliability of well defined essays is greater than that of traditional essays. Marking scheme significantly increases mean assessment, but does not increase any reliability.

Limitations and Suggestions :

At least in this investigation, the marking scheme devised by the investigator did not prove effective. On the basis of the discussion with the examiners some reasons regarding the ineffectiveness of the marking scheme are given.

There were differences of opinions among the examiners regarding the different aspects of written expression. These differences were not cleared off as a result of which inter-examiner reliability might have been decreased. This defect could be over come by briefing the examiners before

the work of assessment is carried out.

Specification regarding the weightages to be given to the spellings and corrections of the language was faulty. Deduction of marks was to be carried out on the basis of number of mistakes committed. So the students who wrote a very small amount of content, gave a little chance ~~to~~ of deduction of marks while in comparison with this, those who wrote a substantial amount of content, comitted more mistakes and were penalised, which was not a fair thing for good discrimination. Such loop-holes in the marking scheme also might have decreased the reliability. It would have been better if the marks were deducted on the basis of number of mistakes made in proportion to the amount of written material.

The same marking scheme was applied to all the types of essays, reflective, descriptive, autobiographical etc. The judgements would have been finer if the peculiarities of various categories of essays were considered by preparing separate marking scheme for each category of essay.

The marking scheme was not constructed according to evaluation criteria as the pilot form of the marking scheme was not scrutinised and criticised by the experts for finding out any loop-holes and defects, and hence the

suggestions for modifying the marking scheme were not available. It is desirable to repeat this experiment with marking scheme modified by the experts, to get more valid results.

The examiners were required to assess the essays of different categories in the same lotm i.e. after examining an autobiography, the next essay might be of descriptive nature and the third in suggestion might be based on reflictive thinking. Thus there remains every chance of fluctuation the judgement of the examiner might become faulty. For the purpose of this investigation this factor regarding category of the essay should have been controlled, but any how it was not done so. Perhaps an investigation with only one essay (without giving any option to the student) might give better results.

For the purpose of investigation a sample of only 83 students was taken which is of course not very insufficient, but a bigger sample than this (about 400 students) is desirable for getting more stable and valid results.

APPENDIX A

The Two Essay Tests

(Translated)

New Arts College

Essay Test

December 1966 — 1½ Hours

Note : Write both the Answers in separate answer-books

Section I

1. Write an essay on any one : 20

(1) I do not want to soar in Dreams +

(2) A quarrel some Family

(3) Mr. Bhadrambhadra's* visit to a modern college

(4) Co-Education

(5) An Autobiography of a Policeman

+ A quotation from Gujarati verse

* Bhadrambhadra is a quixotic character famous in Gujarati literature.

Section II

2. Write an essay on any one : 20

(1) LO! The soft Sunshine and the south wind invite me!

(2) Description of a castle having dungeons, trenches, secret passages

(3) Ramal went to buy vegetables. He saw there two instances of honesty and two instances of dishonesty. A narration of all this.

(4) Unity of Gur Nation (including the propriety or non-propriety of the following remedies : Abolition of different states, President's Rule, Inter-Communal and Inter-Provincial Marriages)

(5) Through influence you got the job of a peon in an Institution. After eight years you rose to the post of its head-clerk. (Autobiography of a peon)

APPENDIX B

Sample of the front page of the answer-book

New Arts College

Vallabh Vidyanagar

Subject : Gujarati Essay Test Date _____

Section

Marks obtained

Note : Do not write any thing on this page as well as on the back side of this page. Essays from each section are to written in separate answer books.

Write you roll number only once at the bottom, in the extreme right hand corner square.

Roll No. Roll No. Roll No.

Roll No. Roll No. Roll No.

Roll No. Roll No. Roll No.

APPENDIX C

Marking Scheme
(No scheme was given)
Marking Scheme 2

A new thing is proposed here. All the four mental aspects (noedéd in writing an essay) will be evaluated separately, their marks will be assigned separately and then added up to form the total out of 20.

Key

Please assign the marks separately for each aspect, and then add them up. While assigning marks also consider the sub-aspects separately. First add them up aspect-wise.

ASPECTS :

I Material/Thought/Arguments/Imaginative/Power/Narration etc. Max. Marks 6.

A.	Material	3½ Marks
B.	Fund of knowledge and well versedness	½ Marks
C.	Originality	1 Mark
D.	Sense of Propriety	½ Mark
E.	Quotations	½ Mark

II	Presentation		Max Marks 4
	A.	Paragraphing	1 Mark
	B.	Godd Beginning	½ Mark
	C.	Good Ending	½ Mark
	D	Artistic Development of the Material	3 Marks
	E.	Logicality and Logical Sequence	1 Mark
III	Powor of Language Consider in		Max Marks 4
		Toto	4 Marks
IV	Correctness of Language		Max. Marks 4
	A.	Correctness of sentence Formation	1 Mark
	B.	The remaining elements of Grammar	1 Mark
	C.	Spelling	1 Mark
	D.	Punctuation Marks	1 Mark

How to consider the marking of sub-points?

Find details of assigning marks:-

1. A. For Material ½ per point. The rest is clear.
2. A,B Assign marks only if noteworthy.

C. We have to evaluate here the art of presentation and similar creative faculties. If the material is developed in the best manner____3 Marks

If the development is mediocre — 1½ Marks

If there is no head or tail — 0 Mark

D. If the paragraphing is correct on the whole --- — 1 Mark

If there is no head or tail — 0 Mark

If in - between — ½ Mark

E. In C the emphasis was on the artistry of presentation here it is no logicity in presentation. Cut out marks only if there is a discount in logicality.

3. Assign marks in toto.

4. A More than four wrong formations — 0 Mark

Two to three wrong formations — ½ Mark

One or zero wrong formation — 1 Mark

B Four mistakes or more — 0 Mark

Two to three mistakes — ½ Mark

One or zero mistakes — 1 Mark

C False of 6 words or more — 0 Mark

False of 3 to 5 words — ½ Mark

One or two mistakes — 1 Mark

(While deducting marks further repeatition of the same mistakes is to be ignored).

D	A general lack of punctuations	0 Mark
	Some correct punctuations, some incorrect or lacking	$\frac{1}{2}$ Mark
	Nearly all used correctly	1 Mark

APPENDIX D

TABLES

TABLE I

Frequency distribution of differences between the marks of first and second assessments without using the marking scheme

Difference of marks D	Frequencies f							
	TiA	TiB	TiC	TiD	WiA	WiB	WiC	WiD
+ 5	5	-	-	-	1	-	-	-
+ 4	10	-	1	1	2	2	-	5
+ 3	11	4	3	2	2	4	2	8
+ 2	15	3	10	6	12	12	7	13
+ 1	16	13	14	17	20	20	20	20
0	13	31	26	33	32	28	31	18
- 1	5	21	22	16	9	12	11	11
- 2	3	7	6	6	4	5	2	6
- 3	2	3	1	-	1	-	-	1
- 4	1	1	-	2	-	-	-	-
- 5	2	-	-	-	-	-	-	1
Total	83	83	83	83	83	83	83	83
Mean of absolute difference	2.1	0.81	1.0	0.81	1.0	0.77	0.82	1.45

TABLE II

Frequency distribution of differences between the marks of first and second assessments with using the marking scheme

Difference of marks	Frequencies							
	TJA	TJB	TJC	TJD	WJA	WJB	WJC	WJD
+ 5	-	-	-	2	-	-	-	-
+ 4	-	-	3	2	-	-	2	-
+ 3	4	1	4	2	2	1	1	1
+ 2	12	6	12	6	1	1	6	2
+ 1	17	27	9	17	7	3	13	9
0	23	30	10	31	14	14	15	19
- 1	14	15	16	16	21	27	24	26
- 2	9	4	9	5	16	26	10	17
- 3	4	-	10	2	15	9	7	5
- 4	-	-	-	-	3	2	4	3
- 5	-	-	-	-	4	1	1	-
Total	83	83	83	83	83	83	83	83
Mean of absolute difference	1.2	0.8	1.6	1.2	1.74	1.5	1.5	1.2

TABLE III

Frequency distribution of maximum differences between the sets of marks given by the four examiners

Maximum difference	Frequency			
	Ti	TJ	Wi	WJ
1	3	1	5	1
2	8	7	19	1
3	20	10	24	17
4	27	15	20	20
5	18	23	12	20
6	4	16	1	17
7	-	7	2	6
8	2	3	-	1
9	1	1	-	-
Total	83	83	83	83
Average of the maximum difference	3.81	4.78	3.31	4.65

TABLE IV

Matrix of Assessment of Traditional Essays without the marking scheme

Sr. No.	Roll No.	TiA1	TiA2	TiB1	TiB2	TiC1	TiC2	TiD1	TiD2
1	6	10	14	8	8	6	8	10	10
2	9	12	9	9	8	8	7	10	10
3	15	7	8	7	7	9	8	9	9
4	19	3	3	6	3	2	3	8	4
5	22	7	8	9	8	10	8	11	11
6	27	7	8	9	8	11	10	9	11
7	30	7	7	7	7	6	8	9	9
8	33	4	6	7	7	8	6	9	8
9	35	7	10	11	11	10	9	10	12
10	37	8	10	9	9	9	9	10	10
11	41	8	7	8	8	7	7	10	11
12	42	9	4	7	0	7	4	9	2
13	44	6	8	8	8	8	7	9	9
14	49	8	6	7	8	9	8	9	9
15	55	4	4	9	7	7	6	5	8
16	56	9	12	10	8	7	6	9	10
17	57	6	5	10	8	8	7	8	9
18	60	4	7	5	3	4	4	6	4
19	61	4	5	5	5	5	5	5	8

contd...

Table IV contd.

Sr. No.	Roll No.	TiA1	TiA2	TiB1	TiB2	TiC1	TiC2	TiD1	TiD2
20	62	8	9	8	8	8	9	10	9
21	63	7	9	8	7	6	5	9	10
22	69	7	9	7	7	8	8	9	10
23	75	5	9	9	8	6	5	9	9
24	78	7	11	11	8	8	6	10	9
25	81	5	10	7	8	8	6	9	10
26	83	7	11	7	7	10	8	8	8
27	86	8	16	9	8	10	10	10	12
28	87	7	16	12	10	10	10	10	10
29	90	7	9	8	8	10	8	10	9
30	93	7	10	8	7	9	8	11	10
31	95	4	5	3	2	0	2	2	3
32	96	3	3	4	7	2	5	3	2
33	97	8	11	8	8	12	12	10	11
34	98	6	8	7	7	11	12	11	10
35	99	9	13	8	9	9	10	9	10
36	100	8	8	7	10	12	12	10	10
37	102	6	6	6	8	7	7	9	8
38	105	1	3	2	1	1	1	2	2
39	106	5	8	7	8	7	7	9	8
40	108	7	8	8	9	9	8	10	10
41	109	4	7	6	8	7	7	8	8

contd...

Table IV Contd..

Sr. No.	Roll No.	TiA1	TiA2	TiB1	TiB2	TiC1	TiC2	TiD1	TiD2
42	109B	7	10	9	9	7	6	8	9
43	144	5	7	7	6	8	8	6	8
44	166	0	0	0	0	1	1	1	1
45	180	3	6	6	5	6	6	7	6
46	181	4	5	6	6	7	7	8	7
47	213	7	8	7	8	9	9	10	10
48	214	6	5	8	7	7	7	9	8
49	225	6	6	7	7	6	8	8	8
50	227	7	8	7	7	10	8	10	10
51	229	5	6	7	3	6	5	8	6
52	232	7	5	7	7	7	6	9	9
53	233	5	5	7	7	7	7	6	6
54	235	4	6	7	8	7	6	6	7
55	239	11	10	11	10	13	14	10	10
56	243	6	10	10	9	12	10	9	11
57	245	6	11	7	7	8	7	10	10
58	249	8	8	10	10	6	8	11	11
59	252	5	7	7	7	9	10	10	9
60	255	7	9	8	7	8	11	9	10
61	256	7	12	7	9	8	12	10	10
62	257	7	10	8	9	9	11	6	10
63	258	7	5	3	3	4	3	6	5

contd...

Table IV Contd.

Sr. No.	Roll No.	TiA1	TiA2	TiB1	TiB2	TiC1	TiC2	TiD1	TiD2
64	309	11	7	9	7	8	8	10	6
65	314	4	5	0	3	4	6	8	8
66	318	4	4	6	4	5	4	4	2
67	371	8	9	8	8	12	10	10	10
68	372	5	7	7	8	8	7	9	7
69	384	5	7	6	6	7	6	8	8
70	392	8	12	11	8	8	8	10	9
71	394	7	7	9	8	8	8	3	8
72	401	9	16	8	8	7	9	11	11
73	402	10	15	8	8	7	9	11	12
74	404	5	8	6	7	6	7	9	8
75	413	2	3	4	4	2	2	2	2
76	545	5	10	6	7	7	8	8	7
77	575	6	9	7	8	8	8	9	8
78	595	6	7	8	8	9	11	10	10
79	639	8	3	8	7	6	6	2	3
80	644	7	8	8	8	10	13	8	8
81	659	7	7	8	7	7	7	9	9
82	666	5	7	8	10	4	5	10	8
83	738	7	11	8	8	9	10	10	8

TABLE V

Matrix of assessment of Traditional Essays with Marking Scheme

Sr. No.	Roll No	TJA1	TJA2	TJB1	TJB2	TJC1	TJC2	TJD1	TJD2
1	6	9	12	7	9	12	13	12	11
2	9	9	9	7	8	11	9	12	10
3	15	11	10	8	7	11	12	9	9
4	19	7	6	4	4	8	7	7	5
5	22	6	7	7	8	12	13	11	11
6	27	8	6	7	8	13	12	11	11
7	30	7	8	6	7	13	11	10	9
8	33	6	5	7	8	10	8	8	8
9	35	8	8	8	10	13	12	12	12
10	37	11	12	7	8	10	11	11	12
11	41	8	11	8	8	8	11	10	8
12	42	5	6	2	3	6	7	4	4
13	44	9	7	7	7	10	10	11	10
14	49	7	6	7	7	11	8	10	11
15	55	5	7	8	6	10	10	9	9
16	56	15	12	7	10	7	8	10	12
17	57	8	7	8	9	14	11	10	12
18	60	7	9	3	3	8	7	6	4
19	61	5	6	5	4	5	5	7	7
20	62	8	8	7	8	12	13	9	9

contd..

Table V contd..

Sr. No.	Roll No.	TJA1	TJA2	TJB1	TJB2	TJC1	TJC2	TJD1	TJD2
21	63	7	7	7	7	8	9	10	9
22	69	8	8	7	6	8	10	10	9
23	75	11	11	8	8	7	5	11	11
24	78	9	6	9	8	14	11	7	11
25	81	9	7	8	7	7	11	10	10
26	83	10	9	7	6	13	11	10	7
27	86	10	7	8	9	11	12	11	12
28	87	8	10	7	8	13	11	11	10
29	90	9	6	8	7	8	11	10	12
30	93	10	8	8	8	9	11	11	11
31	95	7	9	3	3	6	7	4	4
32	96	3	3	6	8	7	8	6	10
33	97	10	10	7	7	14	13	11	10
34	98	9	8	8	9	11	11	11	10
35	99	9	10	8	8	12	11	10	11
36	100	10	12	10	9	14	13	10	10
37	102	11	11	6	7	10	13	8	9
38	105	7	7	2	2	3	5	2	3
39	106	9	9	7	7	11	12	9	8
40	108	8	9	8	9	12	14	12	14
41	109	7	7	6	7	8	10	7	6
42	109B	10	12	6	7	9	8	10	10

contd...

Table V Contd..

Sr. No.	Roll No.	THA1	TJA2	TJB1	TJB2	TJC1	TJC2	TJD1	TJD2
43	144	10	11	8	8	11	10	9	8
44	166	4	5	1	1	4	5	2	2
45	180	7	7	6	8	8	8	9	7
46	181	7	8	6	5	8	11	7	7
47	213	8	9	7	7	12	12	10	10
48	214	7	6	7	8	13	11	10	10
49	225	8	6	6	7	12	9	9	9
50	227	8	8	8	8	11	9	10	13
51	229	7	10	6	7	5	6	9	10
52	232	6	7	7	6	11	8	10	10
53	233	7	6	7	7	6	8	9	6
54	235	4	6	9	9	9	11	10	10
55	239	12	10	8	8	15	13	11	11
56	243	13	11	10	9	12	11	11	11
57	245	9	9	8	6	12	9	10	11
58	249	7	6	9	10	12	13	12	12
59	252	7	6	7	7	14	11	10	9
60	255	8	10	8	8	13	10	11	10
61	256	11	12	8	8	12	13	10	10
62	258	9	11	7	7	13	12	7	6
63	258	8	8	5	3	11	8	6	4
64	309	8	10	7	8	10	11	12	11

contd...

Table V Contd..

Sr. No.	Roll No.	TJA1	TJA2	TJB1	TJB2	TJC1	TJC2	TJD1	TJD2
65	314	7	8	7	7	5	5	8	10
66	318	5	3	5	4	6	10	10	5
67	371	12	12	9	7	12	13	11	12
68	372	8	8	7	8	13	13	7	7
69	384	8	6	6	7	5	11	8	8
70	392	9	7	8	9	12	11	10	= 10
71	394	5	5	8	8	14	11	7	12
72	401	11	12	7	7	9	8	11	12
73	402	10	9	8	8	12	11	12	12
74	404	7	9	6	7	10	9	8	9
75	413	6	4	4	3	8	7	3	4
76	545	9	10	6	5	10	10	8	9
77	575	7	8	6	6	7	11	9	8
78	595	8	9	7	7	10	11	10	11
79	639	7	7	6	6	6	8	5	8
80	644	8	11	6	8	12	12	11	10
81	659	5	5	6	7	9	12	9	10
82	666	7	7	8	9	7	9	10	12
83	738	9	9	9	8	9	11	10	10

TABLE VI

Matrix of assessment of Well defined essays without marking scheme

Sr. No.	Roll No.	WiA1	WiA2	WiB1	WiB2	WiC1	WiC2	WiD1	WiD2
1	6	7	7	8	9	12	8	9	9
2	9	7	8	7	10	9	8	8	9
3	15	1	1	1	1	2	2	2	2
4	19	9	13	9	13	7	9	10	12
5	22	3	3	2	3	5	5	4	4
6	27	5	5	7	8	6	7	5	9
7	30	5	6	7	7	7	7	7	10
8	33	4	6	3	4	5	5	4	4
9	35	7	6	6	9	9	9	9	9
10	37	6	5	5	7	7	7	8	8
11	41	5	5	5	7	6	7	6	9
12	42	6	8	7	9	9	11	9	13
13	44	5	6	6	7	8	7	6	8
14	49	7	7	7	8	9	9	8	9
15	55	7	10	8	9	8	8	9	10
16	56	10	11	9	12	7	7	10	11
17	57	6	7	7	9	9	9	8	9
18	60	1	1	0	0	2	1	1	0
19	61	3	5	5	7	6	7	8	8
20	62	6	6	7	7	8	7	6	6

contd...

Table VI Contd..

Sr. No.	Roll No.	WiA1	WiA2	WiB1	WiB2	WiC1	WiC2	WiD1	WiD2
21	63	3	4	3	4	5	6	3	5
22	69	5	5	6	6	6	7	7	8
23	75	5	5	8	7	4	7	8	11
24	78	8	13	9	10	10	10	8	12
25	81	6	7	8	6	8	6	7	9
26	83	3	5	5	4	6	6	5	5
27	86	1	3	2	0	2	2	2	3
28	87	8	10	7	9	9	9	7	11
29	90	6	7	8	10	10	9	8	10
30	93	6	6	7	9	8	6	8	10
31	95	4	4	3	3	6	6	3	5
32	96	4	6	5	6	7	7	8	8
33	97½	4	8	7	7	7	8	9	9
34	98	5	7	7	7	7	8	9	10
35	99	11	13	8	10	10	9	8	10
36	100	7	9	8	8	8	8	10	10
37	102	6	6	7	7	7	7	10	8
38	105	1	1	0	0	1	1	0	0
39	106	6	6	8	8	8	7	9	8
40	108	6	5	6	5	7	7	9	8
41	109	2	2	3	3	4	4	3	4
42	109B	15	13	6	6	9	10	10	8

Contd..

Table VI Contd.

Sr. No.	Roll No.	WiA1	WiA2	WiB1	WiB2	WiC1	WiC2	WiD1	WiD2
43	144	5	5	7	7	6	6	9	8
44	166	13	13	9	8	9	9	11	10
45	180	4	4	5	5	5	6	6	6
46	181	5	3	4	5	6	6	3	4
47	213	6	8	7	7	8	9	9	9
48	214	7	7	9	9	9	10	7	10
49	225	7	6	5	6	6	6	7	8
50	227	5	4	4	4	5	4	5	5
51	229	2	5	2	4	6	6	3	6
52	232	5	5	6	7	6	7	5	9
53	233	4	4	6	7	6	7	8	8
54	235	7	8	8	7	6	7	8	10
55	239	10	10	6	9	10	10	10	11
56	243	6	5	6	8	8	8	9	8
57	245	8	5	6	8	6	7	8	8
58	249	3	3	3	7	5	4	3	5
59	252	4	3	6	6	6	7	8	6
60	255	7	7	8	9	8	7	9	10
61	256	5	3	5	5	5	6	5	8
62	257	7	8	8	6	7	8	7	8

contd...

Table VI Contd...

Sr. No.	Roll No.	WiA1	WiA2	WiB1	WiB2	WiC1	WiC2	WiD1	WiD2
63	258	7	6	7	6	7	8	9	8
64	309	7	7	10	8	8	10	8	10
65	314	8	11	9	8	10	11	6	13
66	318	4	5	6	6	6	6	6	9
67	371	4	4	6	4	6	6	6	8
68	372	8	6	8	8	9	10	11	9
69	384	7	7	7	8	8	8	8	11
70	392	3	2	2	2	2	4	3	3
71	394	10	8	10	9	8	9	10	11
72	401	6	5	4	5	6	7	6	5
73	402	6	7	9	9	8	10	9	10
74	404	6	6	6	5	7	8	8	9
75	413	3	3	5	5	5	6	9	8
76	545	4	5	6	7	6	8	6	8
77	575	5	5	4	5	4	7	3	4
78	595	5	6	8	7	7	9	8	8
79	639	4	4	6	4	7	6	7	3
80	644	4	3	2	3	3	4	4	3
81	659	4	5	6	6	7	8	7	8
82	666	2	3	1	1	2	2	0	2
83	738	5	5	6	5	6	7	6	6

TABLE VII

Matrix of assessment of well defined essays with marking scheme

Sr. No.	Roll No.	WJA1	WJA2	WJB1	WJB2	WJC1	WJC2	WJD1	WJD2
1	6	7	7	11	8	10	11	10	9
2	9	13	8	11	8	11	11	11	10
3	15	5	2	1	1	3	5	3	2
4	19	9	9	10	9	14	14	12	12
5	22	7	6	4	2	9	8	6	4
6	27	8	8	7	8	11	8	9	8
7	30	8	6	8	6	10	12	10	11
8	33	7	8	4	2	5	7	5	4
9	35	9	7	10	9	12	13	10	10
10	37	8	7	7	6	9	10	6	6
11	41	10	7	7	9	11	11	10	9
12	42	11	8	9	8	12	13	13	11
13	44	9	6	7	6	9	9	7	6
14	49	9	7	8	7	12	10	9	9
15	55	6	9	7	7	10	11	10	11
16	56	14	9	9	8	13	13	12	11
17	57	6	4	8	8	13	11	11	9
18	60	3	1	0	1	1	2	1	1
19	61	7	4	6	4	11	9	7	7
20	62	9	8	8	8	9	9	7	5
21	63	7	7	5	4	8	8	5	4

contd..

Table VII Contd...

Sr. No.	Roll No.	WJA1	WJA2	WJB1	WJB2	WJC1	WJC2	WJD1	WJD2
22	69	8	6	6	5	12	8	7	6
23	75	10	9	9	7	9	10	11	9
24	78	10	9	10	8	12	12	12	12
25	81	10	9	8	7	11	10	9	10
26	83	8	8	4	4	7	9	4	4
27	86	7	6	4	3	6	6	2	3
28	87	11	9	7	7	9	11	10	10
29	90	13	9	8	7	13	11	11	9
30	93	12	9	8	6	11	9	10	10
31	95	8	6	4	2	7	6	5	3
32	96	7	4	6	4	7	11	8	6
33	97	9	8	7	6	12	11	10	9
34	98	11	9	9	6	12	9	10	10
35	99	15	10	7	8	11	10	8	9
36	100	13	13	9	6	14	10	10	9
37	102	11	10	9	6	11	11	9	6
38	105	4	0	0	0	2	1	1	1
39	106	9	7	6	5	10	9	8	6
40	108	10	7	6	4	11	10	10	6
41	109	8	7	5	3	6	8	6	3
42	109B	11	13	6	5	13	9	11	8

contd....

Table VII Contd..

Sr. No.	Roll No.	WJA1	WJA2	WJB1	WJB2	WJC1	WJC2	WJD1	WJD2
43	144	8	9	9	7	13	8	11	6
44	166	10	11	9	7	14	12	12	10
45	180	7	7	6	5	10	8	6	5
46	181	8	7	5	4	11	8	5	4
47	213	12	8	8	8	11	10	9	8
48	214	9	9	7	7	13	12	10	10
49	225	6	5	8	4	9	8	6	6
50	227	6	6	5	3	8	8	6	5
51	229	6	6	5	4	9	8	6	6
52	232	4	7	6	6	9	10	7	6
53	233	7	6	6	4	12	9	6	5
54	235	13	11	9	7	12	13	12	10
55	239	11	10	9	8	12	11	11	10
56	243	9	9	8	4	9	8	8	6
57	245	11	9	8	3	9	8	7	5
58	249	5	5	4	3	8	7	4	3
59	250	7	6	6	5	10	7	8	4
60	255	11	9	7	5	10	7	11	11
61	256	11	8	7	6	9	10	9	7
62	257	7	6	8	6	9	11	9	8
63	258	11	8	7	5	12	10	8	7

Contd.....

Table VII Contd..

Sr. No.	Roll No.	WJA1	WJA2	WJB1	WJB2	WJC1	WJC2	WJD1	WJD2
64	309	8	9	9	7	12	11	10	10
65	314	7	8	7	7	13	12	12	12
66	318	6	6	7	6	8	9	6	9
67	371	8	6	7	5	8	9	9	6
68	372	13	11	9	8	14	13	10	9
69	384	7	6	9	6	12	11	11	9
70	392	6	7	2	2	4	6	3	3
71	394	9	6	9	7	12	11	10	11
72	401	7	6	7	5	10	10	6	4
73	402	7	7	10	7	14	13	9	11
74	404	7	7	6	6	9	9	6	7
75	413	8	9	5	4	10	8	8	6
76	545	10	7	4	7	10	9	8	5
77	575	6	5	4	7	10	7	5	5
78	595	9	6	6	5	9	9	8	9
79	639	10	7	5	5	9	10	6	5
80	644	9	4	4	2	8	4	5	3
81	659	9	7	6	4	10	9	8	6
82	666	6	7	6	4	10	9	8	6
83	738	9	7	7	7	12	8	5	6

BIBLIOGRAPHY

1. Fisher R.A. — Statistical Methods for research Workers Oliver and Boyd; Edinburg Tweeddole Court, London.

2. Garrett H.E. Ph.D. — Statistics in Psychology and Education : Alied Pacific Private Limited, Bombay.

3. Gayen A.K. Pandya P.B. Magur R.K. Duari P. Dubey S.D. Bhattacharya N. — Report No. 1 of the Research Project on Examinations Sponsored by Ministry of Education, Government of India, New Delhi.

4. Guilford J.P. — Fundamental Statistics in Psychology and Education McGraw-Hill Book Company INC, New York, Toronto, London.

5. Guilford J.P. — Psychometric Methods McGraw-Hill Book Company, INC, New York Toronto, London.

6. Person, E.S. and Hartly H.O. — Biometrica Tables for Statisticians Volume I Cambridge University Press, London.

7. Maxwell A.E. — Analysing Quantitative Data : Dondon : Methuen & Co., New York John Wiley & Sons., Inc.